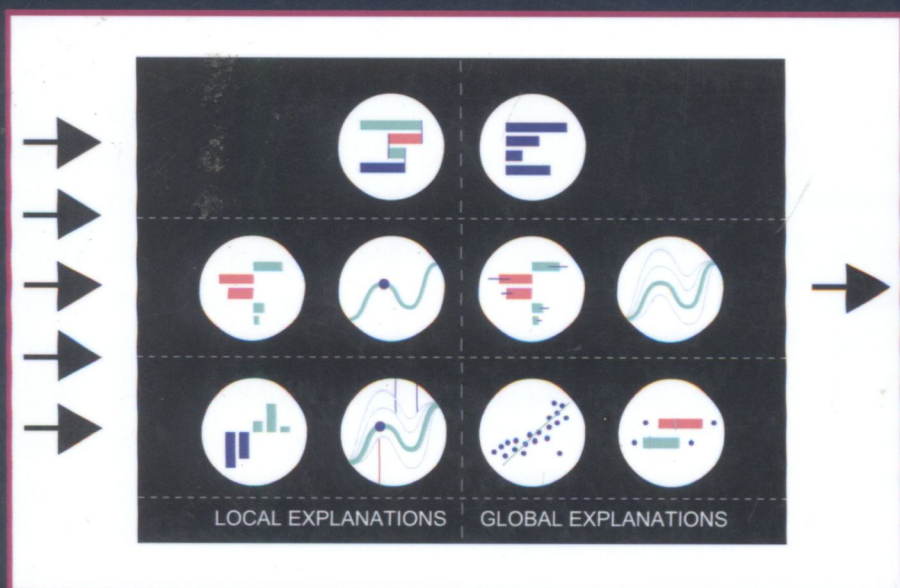


DATA SCIENCE SERIES

EXPLANATORY MODEL ANALYSIS

Explore, Explain, and
Examine Predictive Models



PRZEMYSŁAW BIECEK
TOMASZ BURZYKOWSKI



CRC Press
Taylor & Francis Group

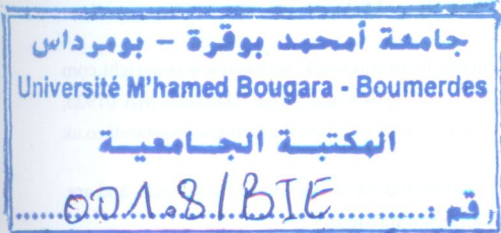
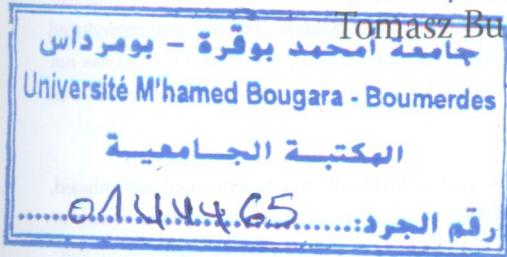
A CHAPMAN & HALL BOOK



Explanatory Model Analysis

Explore, Explain and Examine
Predictive Models

Przemyslaw Biecek
Tomasz Burzykowski



014.44.65



CRC Press

Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Contents

I	Introduction	1
1	Introduction	3
1.1	The aim of the book	3
1.2	A bit of philosophy: three laws of model explanation	6
1.3	Terminology	7
1.4	Black-box models and glass-box models	8
1.5	Model-agnostic and model-specific approach	10
1.6	The structure of the book	11
1.7	What is included in this book and what is not	14
1.8	Acknowledgements	14
2	Model Development	17
2.1	Introduction	17
2.2	Model-development process	18
2.3	Notation	20
2.4	Data understanding	22
2.5	Model assembly (fitting)	24
2.6	Model audit	28
3	Do-it-yourself	29
3.1	Do-it-yourself with R	29
3.1.1	What to install?	29
3.1.2	How to work with DALEX?	30
3.1.3	How to work with <code>archivist</code> ?	30
3.2	Do-it-yourself with Python	31
3.2.1	What to install?	31
3.2.2	How to work with <code>dalex</code> ?	32
3.2.3	Code snippets for Python	32
4	Datasets and Models	35
4.1	Sinking of the RMS Titanic	35
4.1.1	Data exploration	36
4.2	Models for RMS Titanic, snippets for R	40
4.2.1	Logistic regression model	40
4.2.2	Random forest model	40
4.2.3	Gradient boosting model	40

4.2.4	Support vector machine model	41
4.2.5	Models' predictions	41
4.2.6	Models' explainers	43
4.2.7	List of model-objects	45
4.3	Models for RMS Titanic, snippets for Python	46
4.3.1	Logistic regression model	47
4.3.2	Random forest model	47
4.3.3	Gradient boosting model	47
4.3.4	Support vector machine model	48
4.3.5	Models' predictions	48
4.3.6	Models' explainers	49
4.4	Apartment prices	51
4.4.1	Data exploration	52
4.5	Models for apartment prices, snippets for R	54
4.5.1	Linear regression model	54
4.5.2	Random forest model	55
4.5.3	Support vector machine model	55
4.5.4	Models' predictions	55
4.5.5	Models' explainers	56
4.5.6	List of model-objects	57
4.6	Models for apartment prices, snippets for Python	57
4.6.1	Linear regression model	58
4.6.2	Random forest model	58
4.6.3	Support vector machine model	59
4.6.4	Models' predictions	59
4.6.5	Models' explainers	59
II	Instance Level	61
5	Introduction to Instance-level Exploration	63
6	Break-down Plots for Additive Attributions	65
6.1	Introduction	65
6.2	Intuition	65
6.3	Method	69
6.3.1	Break-down for linear models	69
6.3.2	Break-down for a general case	71
6.4	Example: Titanic data	73
6.5	Pros and cons	74
6.6	Code snippets for R	75
6.6.1	Basic use of the <code>predict_parts()</code> function	76
6.6.2	Advanced use of the <code>predict_parts()</code> function	77
6.7	Code snippets for Python	79
7	Break-down Plots for Interactions	83
7.1	Intuition	83

7.2	Method	85
7.3	Example: Titanic data	85
7.4	Pros and cons	88
7.5	Code snippets for R	89
7.6	Code snippets for Python	90
8	Shapley Additive Explanations (SHAP) for Average Attributions	93
8.1	Intuition	93
8.2	Method	94
8.3	Example: Titanic data	97
8.4	Pros and cons	98
8.5	Code snippets for R	99
8.6	Code snippets for Python	102
9	Local Interpretable Model-agnostic Explanations (LIME)	105
9.1	Introduction	105
9.2	Intuition	105
9.3	Method	107
9.3.1	Interpretable data representation	108
9.3.2	Sampling around the instance of interest	109
9.3.3	Fitting the glass-box model	109
9.4	Example: Titanic data	111
9.5	Pros and cons	111
9.6	Code snippets for R	113
9.6.1	The <code>lime</code> package	114
9.6.2	The <code>localModel</code> package	116
9.6.3	The <code>iml</code> package	118
9.7	Code snippets for Python	120
10	Ceteris-paribus Profiles	123
10.1	Introduction	123
10.2	Intuition	123
10.3	Method	124
10.4	Example: Titanic data	124
10.5	Pros and cons	128
10.6	Code snippets for R	129
10.6.1	Basic use of the <code>predict_profile()</code> function	130
10.6.2	Advanced use of the <code>predict_profile()</code> function	131
10.6.3	Comparison of models (champion-challenger)	134
10.7	Code snippets for Python	135
11	Ceteris-paribus Oscillations	139
11.1	Introduction	139
11.2	Intuition	139
11.3	Method	140

11.4	Example: Titanic data	141
11.5	Pros and cons	143
11.6	Code snippets for R	143
11.6.1	Basic use of the <code>predict_parts()</code> function	144
11.6.2	Advanced use of the <code>predict_parts()</code> function	145
11.7	Code snippets for Python	146
12	Local-diagnostics Plots	147
12.1	Introduction	147
12.2	Intuition	147
12.3	Method	149
12.3.1	Nearest neighbors	149
12.3.2	Local-fidelity plot	150
12.3.3	Local-stability plot	151
12.4	Example: Titanic	151
12.5	Pros and cons	152
12.6	Code snippets for R	153
12.7	Code snippets for Python	156
13	Summary of Instance-level Exploration	157
13.1	Introduction	157
13.2	Number of explanatory variables in the model	158
13.2.1	Low to medium number of explanatory variables	159
13.2.2	Medium to a large number of explanatory variables	159
13.2.3	Very large number of explanatory variables	159
13.3	Correlated explanatory variables	160
13.4	Models with interactions	160
13.5	Sparse explanations	161
13.6	Additional uses of model exploration and explanation	161
13.7	Comparison of models (champion-challenger analysis)	162
III	Dataset Level	167
14	Introduction to Dataset-level Exploration	169
15	Model-performance Measures	171
15.1	Introduction	171
15.2	Intuition	172
15.3	Method	173
15.3.1	Continuous dependent variable	173
15.3.1.1	Goodness-of-fit	173
15.3.1.2	Goodness-of-prediction	174
15.3.2	Binary dependent variable	176
15.3.2.1	Goodness-of-fit	176
15.3.2.2	Goodness-of-prediction	177
15.3.3	Categorical dependent variable	181

15.3.3.1	Goodness-of-fit	181
15.3.3.2	Goodness-of-prediction	181
15.3.4	Count dependent variable	182
15.4	Example	183
15.4.1	Apartment prices :	183
15.4.2	Titanic data	183
15.5	Pros and cons	186
15.6	Code snippets for R	187
15.7	Code snippets for Python	189
16	Variable-importance Measures	193
16.1	Introduction	193
16.2	Intuition	194
16.3	Method	194
16.4	Example: Titanic data	195
16.5	Pros and cons	198
16.6	Code snippets for R	198
16.7	Code snippets for Python	204
17	Partial-dependence Profiles	207
17.1	Introduction	207
17.2	Intuition	208
17.3	Method	208
17.3.1	Partial-dependence profiles	208
17.3.2	Clustered partial-dependence profiles	209
17.3.3	Grouped partial-dependence profiles	210
17.3.4	Contrastive partial-dependence profiles	211
17.4	Example: apartment-prices data	212
17.4.1	Partial-dependence profiles	212
17.4.2	Clustered partial-dependence profiles	213
17.4.3	Grouped partial-dependence profiles	214
17.4.4	Contrastive partial-dependence profiles	215
17.5	Pros and cons	216
17.6	Code snippets for R	217
17.6.1	Partial-dependence profiles	217
17.6.2	Clustered partial-dependence profiles	219
17.6.3	Grouped partial-dependence profiles	219
17.6.4	Contrastive partial-dependence profiles	219
17.7	Code snippets for Python	220
17.7.1	Grouped partial-dependence profiles	222
17.7.2	Contrastive partial-dependence profiles	224
18	Local-dependence and Accumulated-local Profiles	227
18.1	Introduction	227
18.2	Intuition	227

18.3	Method	232
18.3.1	Local-dependence profile	232
18.3.2	Accumulated-local profile	234
18.3.3	Dependence profiles for a model with interaction and correlated explanatory variables: an example	235
18.4	Example: apartment-prices data	238
18.5	Pros and cons	239
18.6	Code snippets for R	240
18.7	Code snippets for Python	242
19	Residual-diagnostics Plots	245
19.1	Introduction	245
19.2	Intuition	246
19.3	Method	246
19.4	Example: apartment-prices data	250
19.5	Pros and cons	255
19.6	Code snippets for R	255
19.7	Code snippets for Python	258
20	Summary of Dataset-level Exploration	261
20.1	Introduction	261
20.2	Exploration on training/testing data	262
20.3	Correlated explanatory variables	263
20.4	Comparison of models (champion-challenger analysis)	264
IV	Use-cases	265
21	FIFA 19	267
21.1	Introduction	267
21.2	Data preparation	267
21.2.1	Code snippets for R	269
21.2.2	Code snippets for Python	269
21.3	Data understanding	270
21.4	Model assembly	271
21.4.1	Code snippets for R	272
21.4.2	Code snippets for Python	273
21.5	Model audit	273
21.5.1	Code snippets for R	274
21.5.2	Code snippets for Python	275
21.6	Model understanding (dataset-level explanations)	276
21.6.1	Code snippets for R	279
21.6.2	Code snippets for Python	279
21.7	Instance-level explanations	280
21.7.1	Robert Lewandowski	280
21.7.2	Code snippets for R	284
21.7.3	Code snippets for Python	285

21.7.4 CR7	285
21.7.5 Wojciech Szczęsny	287
21.7.6 Lionel Messi	289
22 Reproducibility	291
22.1 Package versions for R	291
22.2 Package versions for Python	293
Bibliography	295
Index	305

EXPLANATORY MODEL ANALYSIS

Explore, Explain, and
Examine Predictive Models

Explanatory Model Analysis Explore, Explain, and Examine Predictive Models is a set of methods and tools designed to build better predictive models and to monitor their behavior in a changing environment. Today, the true bottleneck in predictive modelling is neither the lack of data, nor the lack of computational power, nor inadequate algorithms, nor the lack of flexible models. It is the lack of tools for *model exploration* (extraction of relationships learned by the model), *model explanation* (understanding the key factors influencing model decisions) and *model examination* (identification of model weaknesses and evaluation of model's performance). This book presents a collection of model agnostic methods that may be used for any black-box model together with real-world applications to classification and regression problems.



CRC Press

Taylor & Francis Group
an informa business

www.routledge.com

CRC Press titles are available as eBook editions in a range of digital formats

STATISTICS

ISBN: 978-0-367-69392-3



9 780367 693923

